

ブログ検索エンジン アクセラナビがPostgreSQLから Oracle10gへの移行を決断した理由とは

アクセラテクノロジー株式会社

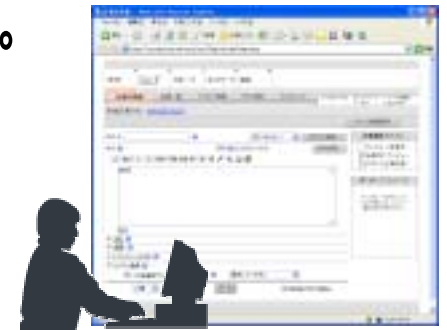
技術本部 萩原純一

2006/3/3

- ブログと、ブログを取り巻く状況
- ブログ検索エンジン アクセラナビのご紹介
- アクセラナビ システムの概要とデータベースの役割
- Oracleデータベースの導入とその効果
- まとめ

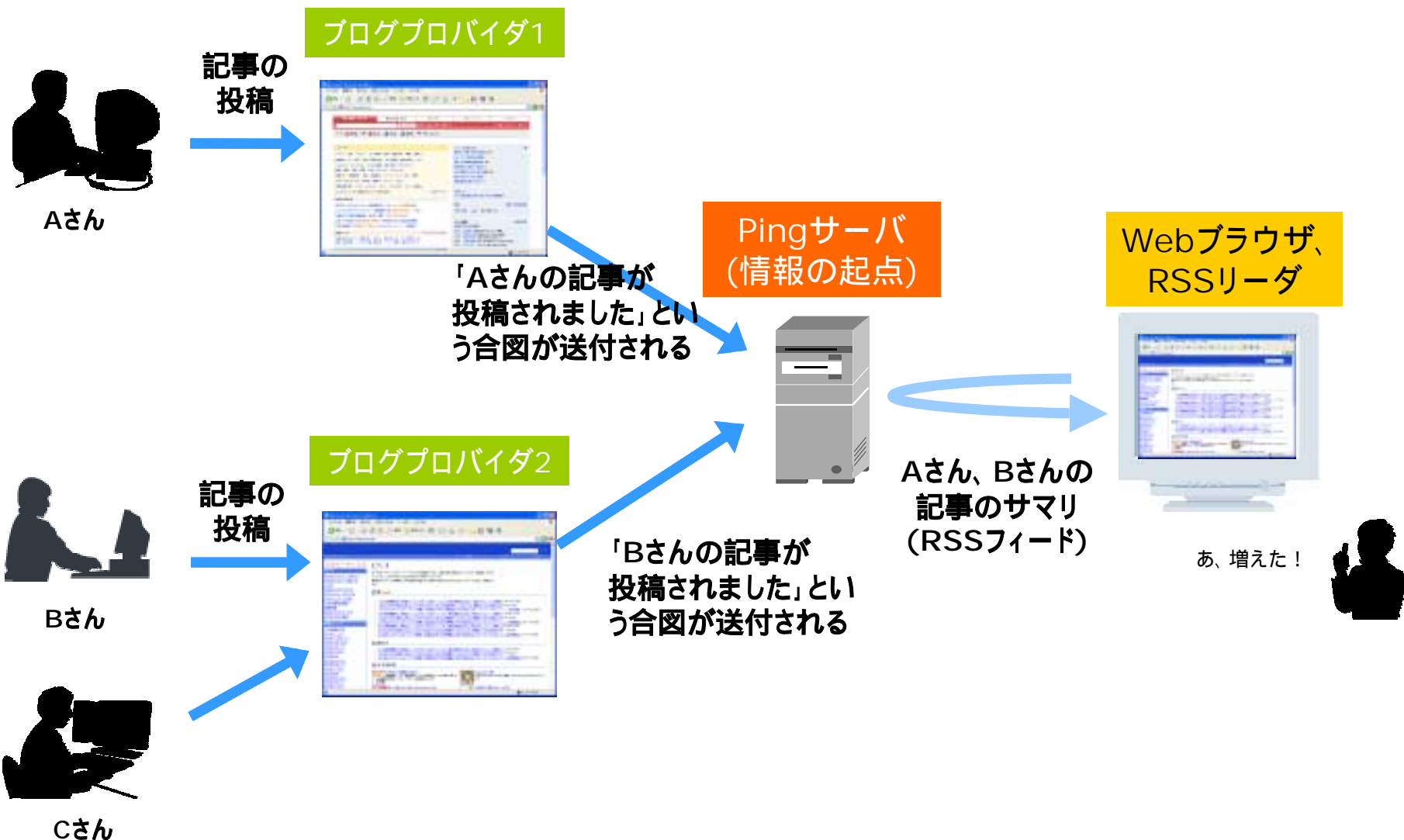
ブログと、ブログを取り巻く状況

- 一般消費者(インターネットユーザ)が日々の出来事や考えていること、商品の使用感などを書きつづけたWebページ
- 従来のWebページによる情報発信と比較すると:
 - 誰でも始められる
大手ポータルサイトがサービスを提供。Webへのアップロードは自動で、書き手はひたすら記事を書けばよい。
 - 即時性がある
「新しい記事を書いた」ことそのものを外部に知らせる手段が用意されており、すぐに読み手に届く。





ブログの仕組み



- **ブログ作成者・閲覧者は急増^(*)**
 - 2005年3月末: 作成者 延べ 約335万人
閲覧者 約1651万人
 - 2007年3月末: 作成者 延べ 約782万人
閲覧者 約3455万人

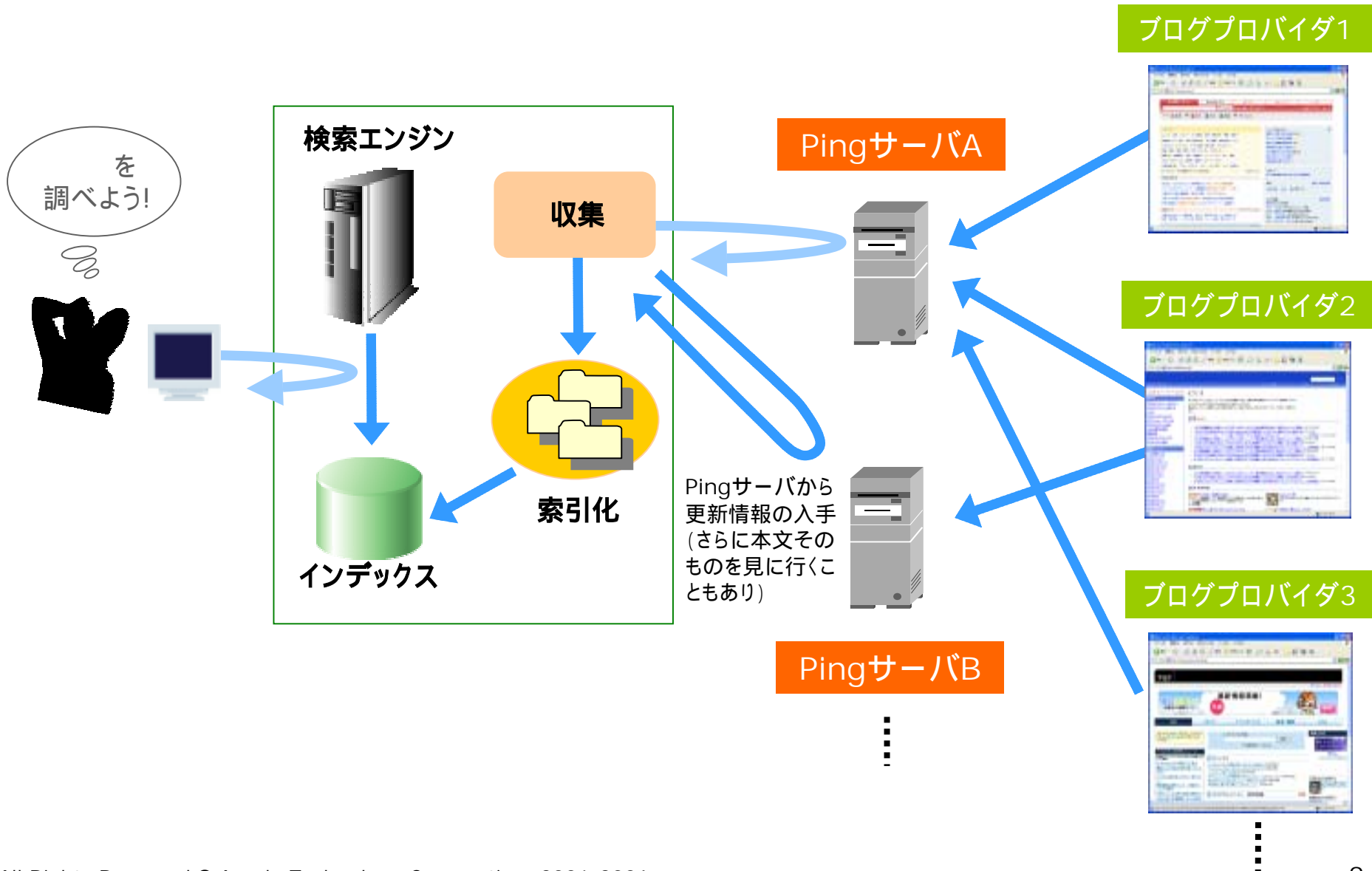
- **一般消費者の生の声としての価値**
 - 企業にとって
マーケティングソースとして活用を模索中。
 - 一般消費者にとって
評判の宝庫。「あの商品の使い勝手は？」など、評判を知りたい。必ずしも、「ブログを見たい」訳ではない。

(*)2005年5月 総務省統計「ブログ・SNSの現状分析及び将来予測」より

ブログだけを対象にした検索サービス

- **高精度な検索**
1つ1つの記事を単位に検索結果を表示。結果として、ゴミの入る確率が低い。Web検索では、記事単位ではなく、ページの単位である。
- **即時性を活かした閲覧**
pingサーバのみを起点に、更新された記事だけをピックアップし、タイムリーに収集・検索対象に出来る。Web検索のように、世の中のWebサイトの全ページをクロールしない。
- **アーカイブとしての価値**
世の中のブログ記事を収集・蓄積。ブログ記事だけを横断的に検索・閲覧できるため、巨大アーカイブとしての価値がある。

ブログ検索システムの仕組み



ブログ検索エンジン アクセラナビのご紹介

■ 設立

2001年7月3日

■ 設立経緯

富士通株式会社からビジネス検索エンジンの開発グループが
スピンアウト。

■ 事業内容

高度ビジネス情報活用ソリューションの提供

- ビジネス向け高性能検索システム(Accela BizSearch)の開発、販売
- 上記システム適用におけるプロフェッショナルサービスの提供

■ 導入実績

プロダクトの販売累計として、約790サーバ(2006年2月現在)

■ ブログ検索専門サイト <http://www.accelanavi.com>



- お客様が快適に使える検索速度
1億超記事を0.1秒～数秒オーダの高速検索
- 日々増加するブログ記事に対する網羅性
ブログ記事の網羅性でトップを目指し(2006/2に1億記事突破)、システム構成・検索エンジンのスケーラビリティを重視
- 欲しい記事を確実に探せる高精度
検索対象はサマリ情報(RSSフィード)だけでなく、いわゆる記事全文。記事中に存在する文字列は必ずヒットし、漏れのない検索を実現
- ブログの即時性を最大限に活かす
しばらく経てば最新記事が検索できる高頻度のインデックス更新を実現
- 純国産エンジンによる安心サポート
検索エンジンとして当社製Accela BizSearchを採用。エンジンから運用まで国内にて迅速かつ柔軟に対応可能

アクセラナビ システムの概要と データベースの役割

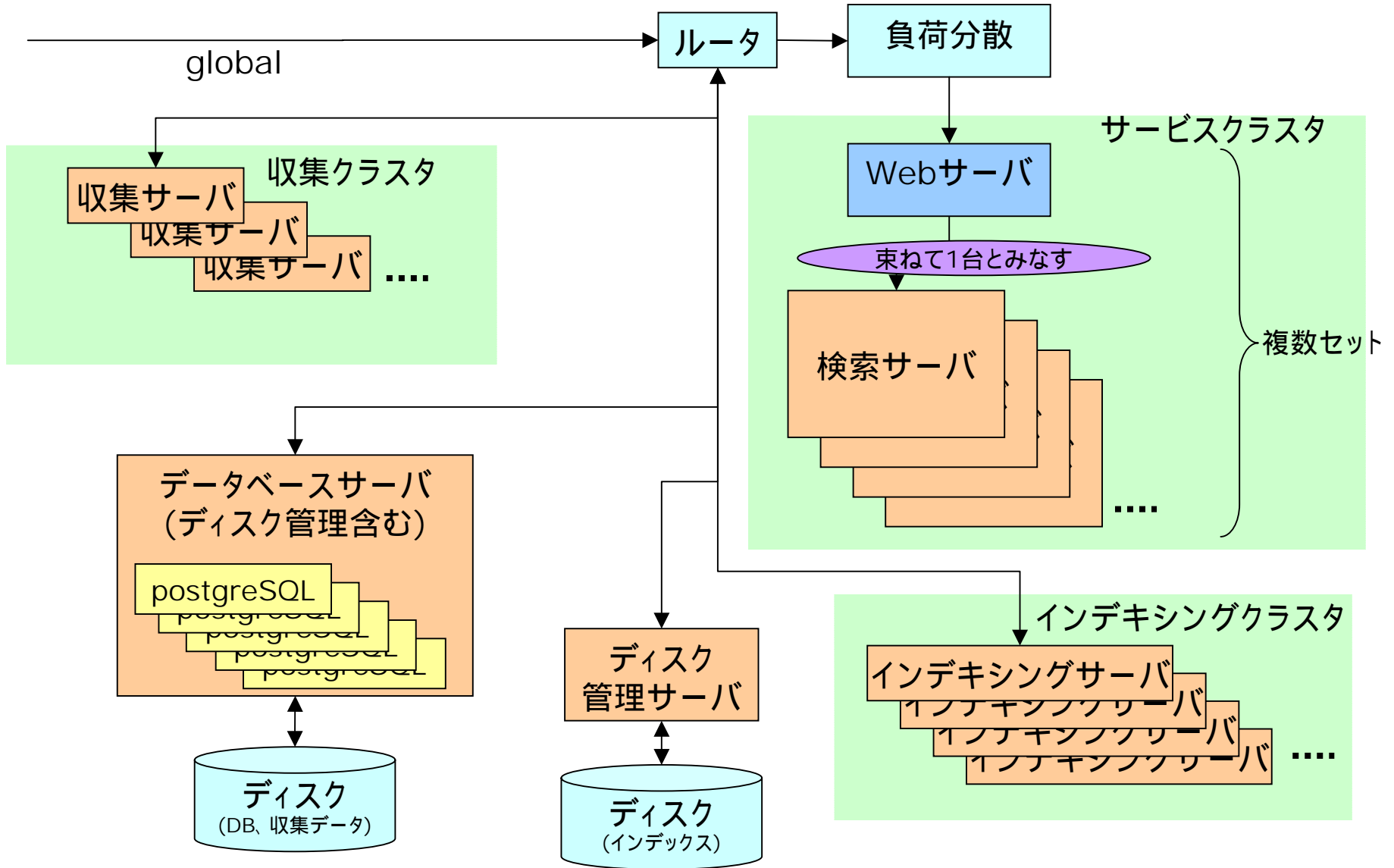
- **検索性能の維持**
ブログ記事数の増加に対して、スケーラブルな検索性能を維持できること
- **更新の即時性**
出来るだけ速く、収集した記事を検索可能にし、サービスへ投入できること

アクセラナビでは、検索専用のソフトウェア (Accela BizSearch) を導入し、上記を実現。

- **収集の高速性**
大量の記事収集が出来ること。今後の記事増加に耐えうること
- **蓄積の効率化**
追加・更新された記事を蓄積すること。特に記事更新時の上書き性能を維持できること

アクセラナビでは、収集部は自作プログラム、蓄積部はデータベース(postgreSQL)を導入し、上記を実現。

アクセラナビシステム概要



- 頻繁な追加・更新に耐えうる安定性
- ブログ記事数の伸びに追従したレコード数への拡張性
- 頻繁な追加・更新に耐えうる「無駄のない」更新方式
- 集まった「価値ある」記事が分析できる機能性

- 大規模なデータベース管理
 - 1日50万記事以上の追加・更新
= 秒間6件以上のsearch & update処理
 - 1億レコード以上の管理
蓄積型サービスであり、減らない。増え続ける
- postgresQLでは...
 - 1億レコードのデータベースを複数インスタンスで分散
= 管理コストが増大中。
 - CPUロードが常に20～40以上！
 - データベースの最適化処理(vacuum)がいつ終わるか
分からない。そもそも終了しない！

- データベースを駆使した分析
 - ブログ作成者のアクティビティ分析
 - ブログ記事そのものの分析、など
- PostgreSQLでは...
 - 高負荷かつ最適化できていない状態であり、分析向けのSELECT文が全く返ってこない
 - 収集以外、何も出来ない

Oracleデータベースの導入とその効果

■ Oracleデータベースの魅力

- スケーラビリティと性能
- 大量データを自動的に分割・管理するパーティショニング機能

■ Miracle Linuxの魅力

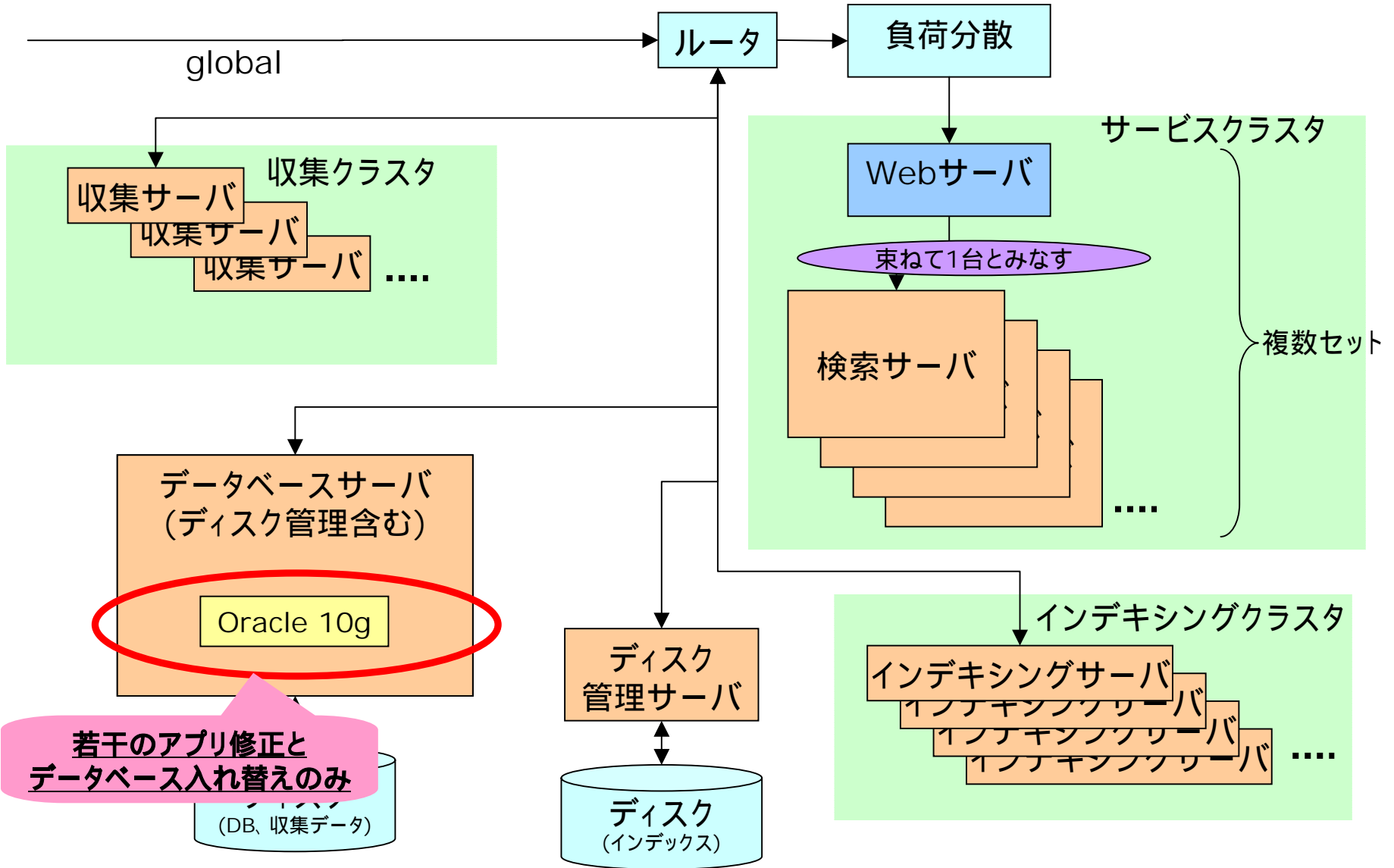
- オープンソース・データベースからの豊富な移行ノウハウ
- OSクラッシュやシステム障害時のサポートと豊富な解析ツール
- 日本語で気軽に質問できるサポート体制
- FAQの充実(特にホームページ)

高度な設計支援

- 当社から、アクセラナビの性能要件、現行/未来のアクセスパターン、現行のpostgresqlのテーブル定義を提示。
- 上記要件を元に、より改善されたOracleのテーブル定義をミラクルリナックス社にて作成。
- 同時に、ミラクルリナックス社から移行予想時間、移行手順を示していただく。

Oracleデータベース初心者の我々としては、
非常に心強いサービス

導入後のシステム構成



大きな安心感: ミラクルリナックス社のサポート

- 何かトラブルのあったときの問い合わせ先として、大きな安心感あり。
- 実際に移行途中において、OS/kernel関連を問い合わせたが、詳細な結果がすぐに返ってくる。
- kernelパラメタに関する知識が豊富で、チューニングのアドバイスを的確に受けられる。

- 収集・蓄積部はOracleにて運用開始
 - 性能: 劇的に向上
 - ロードは1~2程度。
 - 安定性: 1プロセスで運用可能
 - 管理コストが大幅に削減。運用の手間と安心感(対故障の観点)の面で。
 - 拡張性: パラメタ・詳細設定が豊富
 - 詳細な設定が可能。アクセラナビのような性能/安定性重視のシステムでは、チューニングの余地があり、非常によい。
- 収集済みの1億記事はpostgreSQLから移行中

- 収集した記事の分析
 - ブログのアクティビティ調査など、統計としての価値を追求
 - マーケティングインフラとして活用予定
- アクセラナビの性能向上
 - キャッシュ表示において、データベースを有効活用し、検索レスポンスを更に向上できそう
- 全インフラのMiracle Linux化
 - 安定したOS、安心のサポートなど、期待は大きい

まとめ

■ ブログとブログ検索

- ブログは一般消費者の生の声としての価値があり、作成者・閲覧者共に急激に増加していく。
- ブログに特化した検索エンジンは、Web検索ではカバーできない機能性・即時性がポイントとなる。

■ ブログ検索(アクセラナビ)でのデータベース利用

- 合計1億レコード以上、1日50万レコード以上の更新性能が必要。
- 更に、分析等を行えるだけの性能的余裕が必要。
- postgresSQLでは実現できず。

■ Oracle + Miracle Linuxでの解決

- スケーラビリティと性能の確保と、サポート面の安心感が決め手。



Accela

Technology[®]

C O R P O R A T I O N